

Ernst Günter Schukat-Talamazzini

# Automatische Spracherkennung

Grundlagen, statistische Modelle  
und effiziente Algorithmen

Technische Universität Darmstadt  
FACHBEREICH INFORMATIK

B I B L I O T H E K

Inventar-Nr.: 104-00550

Sachgebiete: \_\_\_\_\_

Standort: \_\_\_\_\_



# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
1.1	Mensch-Maschine-Kommunikation . . . . .	1
1.2	Maschinelle Verarbeitung gesprochener Sprache . . . . .	3
1.3	Entwurfsparameter spracherkennender Systeme . . . . .	6
1.4	Warum ist automatische Spracherkennung schwierig? . . . . .	8
1.5	Geschichte und Forschungsparadigmen . . . . .	11
1.6	Wahrscheinlichkeitsorientierte Spracherkennung . . . . .	15
1.7	Übersicht . . . . .	18
<b>2</b>	<b>Gesprochene Sprache</b>	<b>21</b>
2.1	Artikulation und symbolphonetische Beschreibung . . . . .	22
2.1.1	Artikulation von Sprachlauten . . . . .	22
2.1.2	Phonologische Kategorisierung . . . . .	25
2.1.3	Silbe und Intonation . . . . .	28
2.1.4	Ausspracheverschleifung . . . . .	30
2.2	Akustische Theorie der Spracherzeugung . . . . .	32
2.3	Wahrnehmung . . . . .	37
2.3.1	Gehörorgane und Schallwanderung . . . . .	37
2.3.2	Lautheitswahrnehmung . . . . .	40
2.3.3	Frequenzgruppen- und Tonhöhenwahrnehmung . . . . .	41
2.4	Zusammenfassung . . . . .	42
<b>3</b>	<b>Merkmalgewinnung</b>	<b>45</b>
3.1	Diskretisierung . . . . .	46
3.1.1	Abtastung . . . . .	46
3.1.2	Quantisierung . . . . .	47
3.2	Kurzzeitanalyse . . . . .	48
3.2.1	Zeitbereichsmerkmale . . . . .	51
3.2.2	Spektrumanalyse . . . . .	53
3.2.3	Cepstrumkoeffizienten . . . . .	58
3.2.4	Lineare Vorhersage . . . . .	61

3.2.5	Sonstige Ansätze . . . . .	67
3.3	Zeitliche Veränderung des Sprachsignals . . . . .	68
3.4	Zusammenfassung . . . . .	73
<b>4</b>	<b>Klassifikation</b> . . . . .	<b>75</b>
4.1	Numerische Klassifikatoren . . . . .	76
4.1.1	Der optimale Klassifikator . . . . .	76
4.1.2	Parametrische Verteilungsdichtefunktionen . . . . .	79
4.1.3	Verteilungsfreie Klassifikatoren . . . . .	81
4.1.4	Nichtparametrische Klassifikatoren . . . . .	84
4.1.5	Die Schätzung der Fehlerrate . . . . .	86
4.2	Überwachtes Lernen . . . . .	87
4.2.1	Schätzung der Verteilungsdichteparameter . . . . .	87
4.2.2	Informationstheoretische Optimierung . . . . .	90
4.2.3	Verteilungsfreie Verfahren . . . . .	93
4.3	Unüberwachtes Lernen . . . . .	96
4.3.1	Vektorquantisierung . . . . .	97
4.3.2	Identifikation von Mischverteilungen (Teil A) . . . . .	100
4.3.3	Der EM-Algorithmus . . . . .	102
4.3.4	Identifikation von Mischverteilungen (Teil B) . . . . .	104
4.4	Suchverfahren . . . . .	106
4.4.1	Metrische Nächster-Nachbar-Suche . . . . .	108
4.4.2	Nichtmetrische Nächster-Nachbar-Suche . . . . .	110
4.4.3	Sequentielle Klassifikation . . . . .	112
4.5	Merkmaltransformationen . . . . .	113
4.5.1	Karhunen-Loève-Transformation . . . . .	114
4.5.2	Klassenbezogene Transformationen . . . . .	116
4.6	Zusammenfassung . . . . .	118
<b>5</b>	<b>Markovmodelle</b> . . . . .	<b>121</b>
5.1	Das Einzelworterkennungsproblem . . . . .	122
5.2	Markovmodelle . . . . .	125
5.2.1	Definitionen . . . . .	127
5.2.2	Produktionswahrscheinlichkeiten . . . . .	129
5.2.3	Die verborgene Zustandsfolge . . . . .	131
5.2.4	Skalierung und Logarithmierung . . . . .	134
5.3	ML-Schätzung der Modellparameter . . . . .	135
5.3.1	Baum-Welch-Algorithmus . . . . .	136
5.3.2	Viterbi-Training . . . . .	139
5.4	Kontinuierliche Zustandsausgabefunktionen . . . . .	140

5.4.1	Normalverteilungsdichten . . . . .	141
5.4.2	Gaußsche Mischverteilungsdichten . . . . .	142
5.4.3	Semikontinuierliche Markovmodelle . . . . .	144
5.5	Lernstichprobe und Parameterraum . . . . .	145
5.5.1	Mehrfache Modelle — mehrfache Trainingsbeispiele . . . . .	146
5.5.2	Parameterverklebung . . . . .	147
5.5.3	Interpolation . . . . .	149
5.5.4	Glättung . . . . .	151
5.6	Verallgemeinerte Modellkonzepte . . . . .	153
5.6.1	Dauermodellierung . . . . .	153
5.6.2	Korrelierte Ausgabedichten . . . . .	156
5.6.3	Kartesische Produktmodelle . . . . .	158
5.6.4	Weitere Lernverfahren . . . . .	159
5.6.5	Integrierte Merkmalstransformation. . . . .	161
5.7	Zusammenfassung . . . . .	163
<b>6</b>	<b>Akustisch-phonetische Wortmodellierung</b> . . . . .	<b>165</b>
6.1	Wortmodelle . . . . .	167
6.1.1	Die Schätzung von Wortmodellen . . . . .	167
6.1.2	Ganzwortmodelle und Wortuntereinheiten . . . . .	170
6.2	Kontextunabhängige Wortuntereinheiten . . . . .	173
6.2.1	Phonologisch orientierte Wortuntereinheiten . . . . .	174
6.2.2	Hierarchische Wortrepräsentationen . . . . .	176
6.2.3	Akustisch orientierte Wortuntereinheiten . . . . .	178
6.3	Kontextabhängige Wortuntereinheiten . . . . .	179
6.3.1	Phoneme in Kontext . . . . .	180
6.3.2	Modelle kontextabhängiger Phoneme . . . . .	181
6.3.3	Verallgemeinerte Triphone . . . . .	183
6.3.4	Polyphone . . . . .	186
6.3.5	Subphonemische Modellierung . . . . .	187
6.4	Wortgrenzenmodellierung . . . . .	190
6.5	Nichtwörter, unbekannte Wörter und neue Wörter . . . . .	192
6.6	Zusammenfassung . . . . .	196
<b>7</b>	<b>Grammatische Sprachmodelle</b> . . . . .	<b>199</b>
7.1	Linguistisch gesteuerte Spracherkennung . . . . .	200
7.2	Stochastische Grammatiken . . . . .	204
7.2.1	Diskrete stochastische Prozesse . . . . .	204
7.2.2	Äquivalenzklassenbildung . . . . .	207
7.2.3	Faktorisierung und Adaption des Sprachmodells . . . . .	210

7.2.4	Konzeptuelle Grammatiken . . . . .	214
7.3	Schätzung der Sprachmodellparameter . . . . .	215
7.3.1	Glättung relativer Häufigkeiten . . . . .	216
7.3.2	Rückfall auf vergrößerte Statistiken . . . . .	219
7.3.3	Interpolation . . . . .	220
7.3.4	Kategorien . . . . .	225
7.4	Zusammenfassung . . . . .	229
<b>8</b>	<b>Dekodierung kontinuierlicher Sprache</b>	<b>231</b>
8.1	Synchrone Suche . . . . .	233
8.1.1	Netzwerke von Markovmodellen . . . . .	233
8.1.2	Die beste Wortsegmentierung . . . . .	236
8.1.3	Strahlsuche . . . . .	240
8.1.4	Vorwärts-Rückwärts-Suche . . . . .	243
8.2	Asynchrone Suche . . . . .	244
8.2.1	Der A*-Algorithmus . . . . .	245
8.2.2	Die Kellersuche . . . . .	247
8.3	Wortschatzorganisation . . . . .	253
8.4	Mehrphasendekodierung . . . . .	259
8.4.1	Das Prinzip der schrittweisen Verfeinerung . . . . .	259
8.4.2	Die $n$ besten Wortketten . . . . .	261
8.5	Zur Interaktion zwischen akustischem und linguistischem Modell . . . . .	265
8.6	Zusammenfassung . . . . .	268
<b>9</b>	<b>Das ISADORA-System</b>	<b>271</b>
9.1	Rekursive Markovmodelle . . . . .	273
9.1.1	Definition . . . . .	273
9.1.2	Die Vorwärts- und Rückwärtswahrscheinlichkeiten . . . . .	275
9.2	Die Systemarchitektur . . . . .	278
9.2.1	Strukturierte Markovmodelle . . . . .	280
9.2.2	Netzwerkformalismus . . . . .	281
9.2.3	Akustische Modelle . . . . .	283
9.2.4	Reduzible und irreduzible A-Knoten . . . . .	284
9.3	Das Netzwerk zur maschinellen Spracherkennung . . . . .	287
9.3.1	Spracheinheiten unterhalb der Wortebene . . . . .	288
9.3.2	Spracheinheiten innerhalb und oberhalb der Wortebene . . . . .	291
9.3.3	Auswahl expliziter Modelle . . . . .	294
9.3.4	Baumförmige Wortschatzorganisation . . . . .	295
9.4	Lernen . . . . .	297
9.4.1	Initialisierung . . . . .	297

9.4.2	Standardlernverfahren . . . . .	298
9.4.3	Das A.P.I.S.-Lernverfahren für hierarchische Wortmodelle . . . . .	300
9.5	Erkennen . . . . .	306
9.5.1	Rekursiver Viterbi-Algorithmus . . . . .	306
9.5.2	Rückverfolgung . . . . .	308
9.5.3	Geschachtelte symbolische Beschreibungen . . . . .	310
9.5.4	Opake Knoten . . . . .	312
9.6	Zusammenfassung . . . . .	314
<b>10</b>	<b>Experimentelle Untersuchungen</b>	<b>317</b>
10.1	Basiskonfiguration zur Spracherkennung . . . . .	318
10.1.1	Sprachdaten und Erkennungsaufgabe . . . . .	318
10.1.2	Konfiguration des Erkennungssystems . . . . .	319
10.2	Detaillierte Auswertungen zur Spracherkennung . . . . .	321
10.2.1	Sprecherabhängige Worterkennung . . . . .	321
10.2.2	Kodebuchentwurf . . . . .	322
10.2.3	Phonetische Modellierung . . . . .	326
10.2.4	Schätzung linguistischer Sprachmodelle . . . . .	330
10.2.5	Mehrphasendekodierung . . . . .	331
10.3	Erkennung spontan produzierter Äußerungen . . . . .	335
10.3.1	Das Erlanger Bahnauskunftssystem . . . . .	336
10.3.2	Evaluierung spontansprachlichen Datenmaterials . . . . .	338
10.4	Zusammenfassung . . . . .	341
<b>11</b>	<b>Zusammenfassung</b>	<b>343</b>
	<b>Literaturverzeichnis</b>	<b>349</b>
	<b>Sachregister</b>	<b>388</b>
<b>A</b>	<b>Spracherkennung mit ISADORA</b>	<b>395</b>
A.1	Phonetische Umschriftsysteme . . . . .	396
A.2	Phonemische Basiseinheiten . . . . .	397
A.3	Die Kardinalzahlwörter von 1 bis 999 999 . . . . .	398
A.4	Kategoriepaargrammatik für Uhrzeitangaben . . . . .	399
A.5	Evaluierung verschiedener Wortuntereinheiten . . . . .	400
<b>B</b>	<b>Rekursive Markovmodelle</b>	<b>401</b>
B.1	Schätzung der Modellparameter . . . . .	401
B.2	Überführung in ein HMM . . . . .	402